Published in final edited form as:

Biochem Biophys Res Commun. 2004 April 30; 317(2): 401–405. doi:10.1016/j.bbrc.2004.03.068.

An efficient system for small protein expression and refolding

Yuan Cheng* and Dinshaw J. Patel

Structural Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA

Abstract

The low expression yield and poor refolding efficiency of small recombinant proteins expressed in *Escherichia coli* have continued to hinder the large-scale purification of such proteins for structural and biological investigations. A system based on a small fusion partner, the B1 domain of *Streptococcal* protein G (GB1), was utilized to overcome this problem. We have tested this system on a small cysteine-rich toxin, mutant myotoxin α (MyoP20G). The highly expressed fusion protein was refolded using an unfolding/refolding protocol. Due to the small size of GB1, we were able to monitor the unfolding/refolding status by heteronuclear single quantum coherence (HSQC) NMR spectroscopy. The final product yielded well-resolved NMR spectra, with a topology corresponding to the natural product. We conclude that GB1 not only increases the expression level but also enhances the refolding of small proteins.

Keywords

Protein refolding; Protein purification; NMR spectroscopy

Proteins for structural biology are generally generated using fusion protein expression systems, such as glutathione *S*-transferase (GST) and maltose-binding protein (MBP). Recently, a His₆-tagged N-terminal fragment (52 residues) of *Staphylococcal* nuclease R (HR52) has been selected as a smaller efficient fusion partner for the expression of small peptides [1]. This system simplifies the purification protocol due to a one-step affinity purification procedure, and dramatically increases the final yield because of the smaller size of the fusion partner. The use of HR52, which constitutes the hydrophobic core of SNase R, has its limitations, because its high hydrophobicity can interfere with the purification and refolding of hydrophobic peptides. Thus, our attempts to express and purify the small protein myotoxin α (Myo) [2], which has a large proportion of hydrophobic residues, using the HR52 system, resulted in the majority of the fusion protein ending up in inclusion bodies [1]. After cleft from fusion tag, Myo was refolded as described by Nedelkov et al. [3]. Due to the low solubility of Myo, the final product was not sufficient enough to meet NMR studies. To overcome this problem, we have constructed another friendly fusion system for the expression, purification, and refolding of small hydrophobic peptides.

*Corresponding author. Fax: 1-212-717-3066. chengy@mskcc.org (Y. Cheng).

The 56-residue B1 domain of *Streptococcal* protein G (GB1), expressed from *Escherichia coli*, yields a highly soluble and stable protein scaffold, with an exceptionally well-resolved ^1H – ^{15}N HSQC spectrum [4]. Due to its small size, ^1H – ^{15}N HSQC spectra have also been recorded and analyzed for target proteins fused to the GB1 domain [5]. The GB1 domain also enhances the stability (up to sixfold) of fused target proteins, without impacting on their structure [6]. This paper sets out to examine whether the GB1 fusion protein expression system can also impact on the refolding of target proteins, including those that have a high proportion of hydrophobic residues and disulfide linkages.

Myotoxin α is a highly compact 42-residue toxin containing a high proportion of hydrophobic residues and three disulfide bonds. It adopts two interchangeable conformations, due to *cis*–*trans* isomerization at the proline 20 peptide bond. Nevertheless, this equilibrium shifts to a single conformation on mutating proline 20 to glycine (MyoP20G) [3]. Thus, MyoP20G fused to the GB1 domain represents a good test system for the expression and purification, as well as unfolding and refolding, of a small hydrophobic protein containing several disulfide residues. Further, the unfolding–refolding of MyoP20G, when fused to GB1, can be readily monitored using heteronuclear NMR spectroscopy.

Materials and methods

Vector construct

The DNA fragments for His₆-tagged GB1 domain with thrombin digestion site and MyoP20G were generated by overlap-PCR. The single mutation, P20G, of myotoxin α was introduced by changing codon CCG to GGT.

Expression, purification, unfolding, and refolding

A single colony of BL21 (DE3) strains harboring pGBO-MyoP20G vector was cultured in 80 ml Luria–Bertani broth overnight and then used to inoculate 2 liters of ^{13}C , ^{15}N -enriched M9 medium containing 50 mg/ml ampicillin. Cells were grown at 37 °C until an absorbance of 0.8 (600 nm) was obtained prior to the 8 h induction with 0.3 mM IPTG at room temperature. The bacteria were collected and resuspended in 100 ml HS300 buffer (25 mM Hepes, pH 7.5, 300 mM NaCl). After ultrasonication treatment, bacteria were ultracentrifuged at 235,000g for 1 h and the supernatant was loaded onto a 5 ml Hitrap Ni²⁺-chelating column (Pharmacia) that was prepared according to the procedure outlined in the manufacturer's manual and pre-equilibrated with HS300 buffer. The column was extensively washed with HS300 buffer containing 25 mM imidazole. The fusion protein was eluted by HS300 buffer with 250 mM imidazole.

Unfolding of the fusion protein was performed by overnight dialysis against 25 mM Tris buffer, pH 7.5, 5 mM DTT. The sample was next diluted with refolding buffer (25 mM Tris, pH 7.5, 200 mM guanidine chloride, 4 mM GSH, 0.4 mM GSSG, and 0.1% NaN₃) to final concentration of 10 μM . The resulting solution was capped with air and kept at room temperature. The crude refolded product was recollected following passage through a 1 ml Hitrap Ni²⁺-chelating column and purified by the same protocol as outlined above.

The digestion of the fusion protein was done overnight using 200 U of thrombin, while dialyzing against 25 mM Tris buffer, pH 7.5, at 4 °C. Finally, the sample was loaded onto Resource S column (Pharmacia) and then eluted using a NaCl gradient from 0 to 800 mM.

NMR spectra

All NMR spectra with recombinant-expressed and labeled samples were acquired on Varian 600 MHz NMR spectrometers at 20 °C. Samples were prepared by mixing 400 µl of protein solution with 20 µl deuterated water. The triple-resonance experiments used for backbone and side chain resonance assignments included HNCO, HN(CA)CO, CBCA(CO)NH, HNCACB, and HCCH-TOCSY.

Results

pGBO vector construction and fusion protein expression

pGBO (plasmid of GB1 domain fusion expression with thrombin digestion site) was constructed based on pET-28b plasmid (Novagen). The DNA sequence for the His₆-tag and the thrombin digestion site, located between *NcoI* and *NdeI* endonuclease sites, was cut out and replaced with a DNA fragment for His₆-tagged GB1 domain followed by a thrombin digestion sequence (Fig. 1). A linker was inserted between the GB1 domain and the protease digestion site to enhance the cleavage efficiency (Fig. 1B). After thrombin digestion, four residues, GSHM, will be left before the N terminus of target protein.

Purification, unfolding, and refolding

The expressed GB1-MyoP20G fusion protein constituted 30% of the whole cell extract, with 90% of the fusion protein partitioning into the soluble fraction (data not shown). The fusion protein, most likely with reduced cysteine sulfhydryls, due to the reductive environment of the *E. coli* expression system, was next passed through a Ni²⁺-chelating column. The eluted fusion protein gave no observable resonances in an ¹H-¹⁵N HSQC NMR spectrum (data not shown). This could reflect the oxidative characteristics of nickel ions in the chelating column, potentially resulting in an incorrectly folded protein with wrong intra/inter disulfide cross-links, or alternately, aggregate formation given the hydrophobic nature of MyoP20G.

Fortunately, the MyoP20G segment of the fusion protein can be simply unfolded by dialysis against 5 mM DTT. The fusion protein, following dialysis, gave two sets of cross-peaks in the ¹H-¹⁵N HSQC NMR spectrum (Fig. 2A). The majority of the well-dispersed peaks corresponded to those from the fully folded GB1 domain [4], while the remaining well-dispersed peaks together with the poorly dispersed peaks bunched in the center of the spectrum (within dashed ellipse, Fig. 2A) correspond to unfolded MyoP20G.

The fusion protein was refolded following dilution (1 liter) in a refolding buffer (see Materials and methods for details) using a refolding protocol previously reported for the folding of chemically synthesized MyoP20G [3]. The refolding process was monitored on a daily basis by removing 4 ml aliquots, concentrating them 10-fold, and recording ¹H-¹⁵N HSQC NMR spectra. These spectra exhibit peaks corresponding to resonances from unfolded, misfolded, and correctly folded portion of MyoP20G in the fusion protein. The

well-resolved peaks (designated by arrows, Fig. 2B), most likely reflecting the correctly folded MyoP20G, underwent no further change after four days in refolding buffer. Next, the fusion protein in 1 liter of refolding buffer was passed through a small nickel column and treated with thrombin following elution off the column. Basic MyoP20G and acidic GB1 were readily separable on a cation exchange column, with the retained MyoP20G eluted with a salt gradient. The resulting elution profile (Fig. 3A) contained several broad peaks followed by a narrow peak. Fractions from each of the peaks were pooled and concentrated prior to recording ^1H - ^{15}N HSQC NMR spectra. Well-dispersed spectra (Fig. 2C) were only observed for the narrow peak collected from the ion exchange column, consistent with formation of a fully refolded MyoP20G. A fraction from this narrow peak was kept at room temperature for one day and loaded onto the cation-exchange column again. Only one peak was observed (Fig. 3B), indicative of the existence of a single conformation.

The pH of the NMR sample of MyoP20G was next changed from 7.5 (Fig. 2C) to 5.4 (Fig. 2D), which resulted in the observation of additional peaks and an improved signal-to-noise, reflective of slower exchange of amide protons at acidic pH. Finally, all residues (but except Q3) of MyoP20G were assigned in HSQC spectrum (Fig. 2D) based on triple-resonance experiments.

Discussion

The advantage of GB1 over HR52 in our hands is that the majority of the fusion protein (up to 90% for MyoP20G) ends up in the soluble fraction, rather than inclusion bodies. The GB1 system, because of its small size, appears to be ideal for the expression and uniform $^{13}\text{C}/^{15}\text{N}$ -labeling of small peptides (in the 25-mer range), yielding more than 20 mg per liter of M9 culture.

GB1, due to its compact fold and solvent accessibility, impacts neither on the structure nor on refolding of the target protein, but rather increases the solubility of the fusion protein. Furthermore, the stability of GB1 allows a wide range of refolding conditions to be sampled, making it a reliable partner in folding protocols. It should be noted that in the present study the GB1-MyoP20G fusion protein was soluble throughout the entire purification procedure, as well as during the unfolding and refolding experiments. Equally importantly, GB1 lacks cysteine residues, thereby facilitating the refolding environment of cysteine-containing proteins.

The most important aspect of the GB1 system is that it allows monitoring of the refolding pathway by NMR spectroscopy. GB1, due to its small size and high stability, yields strong, well-resolved, and assigned [5] amide signals in the ^1H - ^{15}N HSQC spectrum (Fig. 2A), allowing its resonances to be differentiated from those of the target protein. In the case of refolding of MyoP20G, while as part of the GB1 fusion construct, there is a reduction of resonances of unfolded MyoP20G in the central part of the HSQC spectrum (peaks within dashed ellipsoid in Fig. 2A), along with the appearance of several well-dispersed new peaks (labeled by arrows, Fig. 2B) characteristic of the folded state. MyoP20G has two tryptophans (W32, W34), whose side chain indole NH can be observed in all conditions (boxed areas in Figs. 2A–C and two assigned peaks in Fig. 2D). The movement of these two

cross-peaks from two overlapped peaks in the unfolded state (Fig. 2A) to folded state (Figs. 2B–D) also indicated the higher level refolding.

Our procedure circumvents another issue related to recovery of the refolded protein from the large volumes (1 liter in our case) associated with refolding conditions. Generally, ultra-filtration and dialysis need to be employed prior to chromatographic purification. Fortunately, by keeping the GB1 domain with the target protein during the refolding procedure, the N-terminal His₆-tag is also retained and facilitates the subsequent recovery process. Simply loading the refolding solution onto a Ni-column results in the re-harvesting of >90% of the total fusion protein. The His₆-tag segment is separated from the target protein by GB1, and hence its potential interference with the refolding process is kept to a minimum. During this purification step, well-folded MyoP20G lacking free cysteines will not be oxidized during passage through the Ni²⁺-chelating chromatography procedure.

Four residues, GSHM, are left before the N-terminus of the target protein, following thrombin cleavage of the GB1–target protein junction. It is unlikely that these four relatively neutral residues will significantly impact on the structure and function of the target protein. Indeed, the first three (GSH) residues are not even observed in multi-dimensional NMR spectra due to exchange effects. We have also developed another vector, designated pGBE, in which a Tev protease digestion sequence replaces its thrombin counterpart, thereby providing a more specific digestion site. In this version, only three residues (SHM) are left before the N-terminus of the target protein.

The backbone NMR assignments of refolded ¹³C–¹⁵N-labeled MyoP20G were completed using triple resonance experiments in pH 5.4 buffer, due to the observation of a greater number of cross-peaks at this acidic pH (Fig. 2D), relative to pH 7.5 (Fig. 2C), associated with the refolding protocol. The observation of a single set of backbone NH resonances for MyoP20G in Figs 2C and D supports formation of a single folded species with the correct set of three disulfide cross-links.

Conclusion

The GB1 expression/refolding system has been tested on the MyoP20G protein. The result shows that this system is appropriate for small proteins, especially those that are highly hydrophobic and cysteine-rich in sequence.

References

1. Cheng Y, Liu D, Feng Y, Jing G. An efficient fusion expression system for protein and peptide overexpression in *Escherichia coli* and NMR sample preparation. *Protein Pept. Lett.* 2003; 10:175–181. [PubMed: 12678815]
2. Fox JW, Elzinga M, Tu AT. Amino acid sequence and disulfide bond assignment of myotoxin a isolated from the venom of Prairie rattlesnake (*Crotalus viridis viridis*). *Biochemistry.* 1979; 18:678–684. [PubMed: 570412]
3. Nedelkov D, O'Keefe MP, Chapman TL, Bieber AL. The role of Pro20 in the isomerization of myotoxin a from *Crotalus viridis viridis*: folding and structural characterization of synthetic myotoxin a and its Pro20Gly homolog. *Biochem. Biophys. Res. Commun.* 1997; 241:525–529. [PubMed: 9425304]

4. Gronenborn AM, Clore GM. Rapid screening for structural integrity of expressed proteins by heteronuclear NMR spectroscopy. *Protein Sci.* 1996; 5:174–177. [PubMed: 8771212]
5. Huth JR, Bewley CA, Jackson BM, Hinnebusch AG, Clore GM, Gronenborn AM. Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR. *Protein Sci.* 1997; 6:2359–2364. [PubMed: 9385638]
6. Zhou P, Lugovskoy AA, Wagner G. A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins. *J. Biomol. NMR.* 1997; 20:11–14. [PubMed: 11430750]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

B

MTSHHHHHHTYKL.....TVTENYDIPTSSGLVPR*GSHMYKQC.....KGSG

GB1 (56 aa) Myotoxin P20G (42 aa)

Fig. 1.

Construction of expression vector, pGBO, for expression of MyoP20G. Three restriction endonuclease sites are shown in (A). In the final fusion protein sequence (B), the linker and the thrombin recognition site are revealed by italic characters, with a star denoting the cleavage site.

Fig. 2.

Spectra of samples from different stages of the folding process. (A) ^1H - ^{15}N HSQC spectrum of the fusion protein under unfolding conditions. Sample was in 25 mM Tris buffer, pH 7.5, and 5 mM DTT. The majority of the poorly dispersed signals, which are encircled by a dashed ellipse, correspond to backbone NH resonances of unfolded MyoP20G. (B) ^1H - ^{15}N HSQC of fusion protein during the refolding process. Sample was in refolding buffer solution (see Materials and methods). Arrows indicate signals from NH resonances associated with the folded state of MyoP20G. (C) ^1H - ^{15}N HSQC of MyoP20G after

digestion with protease to remove GB1 and purification. Sample was in 25 mM Tris buffer, pH 7.5, and 460 mM NaCl. (D) ^1H - ^{15}N HSQC of MyoP20G in 25 mM sodium acetate-acetic acid buffer, pH 5.4.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 3. Cation-exchange chromatography profile of refolded MyoP20G. (A) Crude refolding product, (B) purified refolded sample. The narrow peak in (A) corresponding to the correctly folded MyoP20G is labeled by an arrow.